Borrowing Treasures from the Wealth: Deep Transfer Learning through Selective Joint Fine-Tuning

Weifeng Ge Yizhou Yu
The University of Hong Kong

IEEE 2017 Conference on Computer Vision and Pattern Recognition
CVPR July 21-26 HONOLULU 2017

Introduction :

- **The Overfitting Problem of Deep Neural Networks**
  - *Deep neural networks require a large amount of training data.*
  - *Collecting and labeling so much data might be infeasible in many cases.*
- **Feature Extraction and Fine-Tuning**
  - *Off-the-shelf CNN features have been proven to be powerful in a variety of computer vision problems.*
  - *Knowledge learnt from ImageNet / Places / MS COCO can be generalized to other applications.*

Observations :

- **Additional Images Sharing Similar Characteristics with An Existing Training Set Could Strengthen the Kernels in Convolutional Layers**
  - *Additional training images can prevent the convolutional neural networks from overfitting quickly, and help the CNNs to learn much more diversified discriminative features.*
  - *Search additional images using low-level characteristics has less restrictions, which can collect more images than using high-level semantic contents.*

Nearest Neighbor Retrieval:

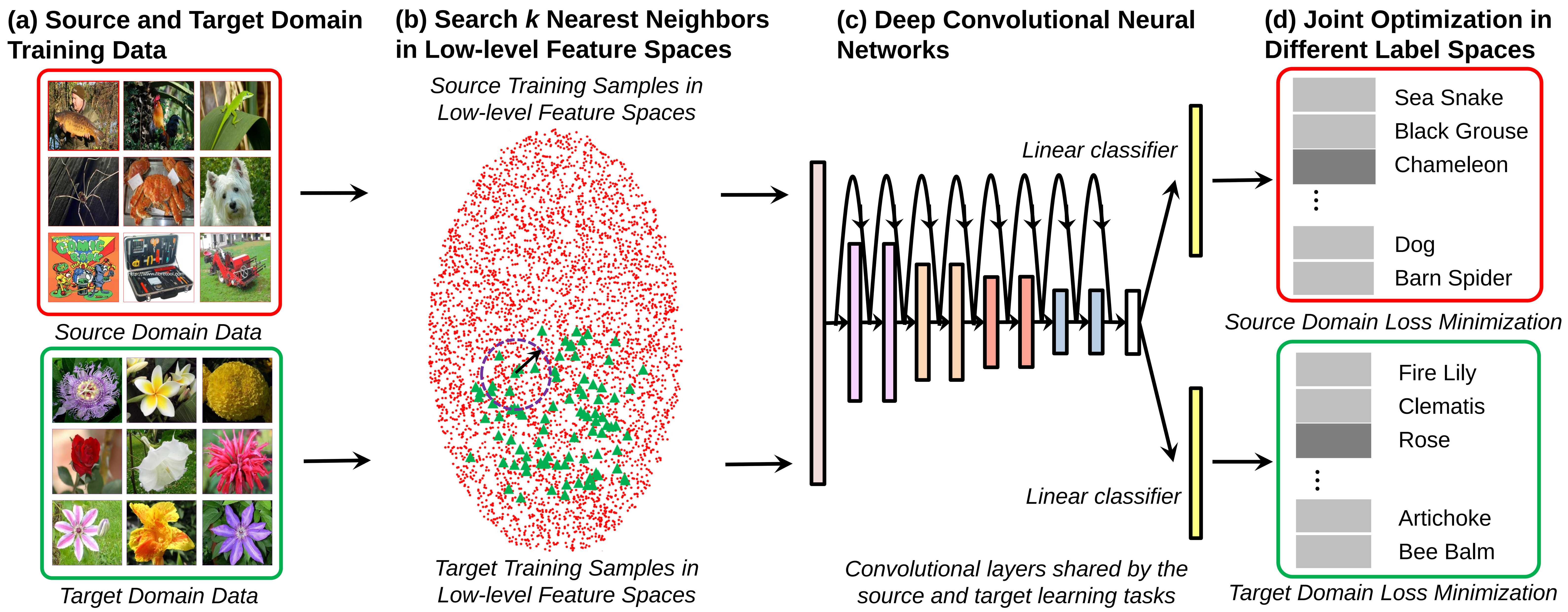
- **Filter Banks (Low-level Feature Response Maps) for Nearest Neighbor Retrieval**
  - *Convolutional Filter Bank*
  - *Gabor Filter Bank*
- **Rank Nearest Neighbors using JS Divergence**

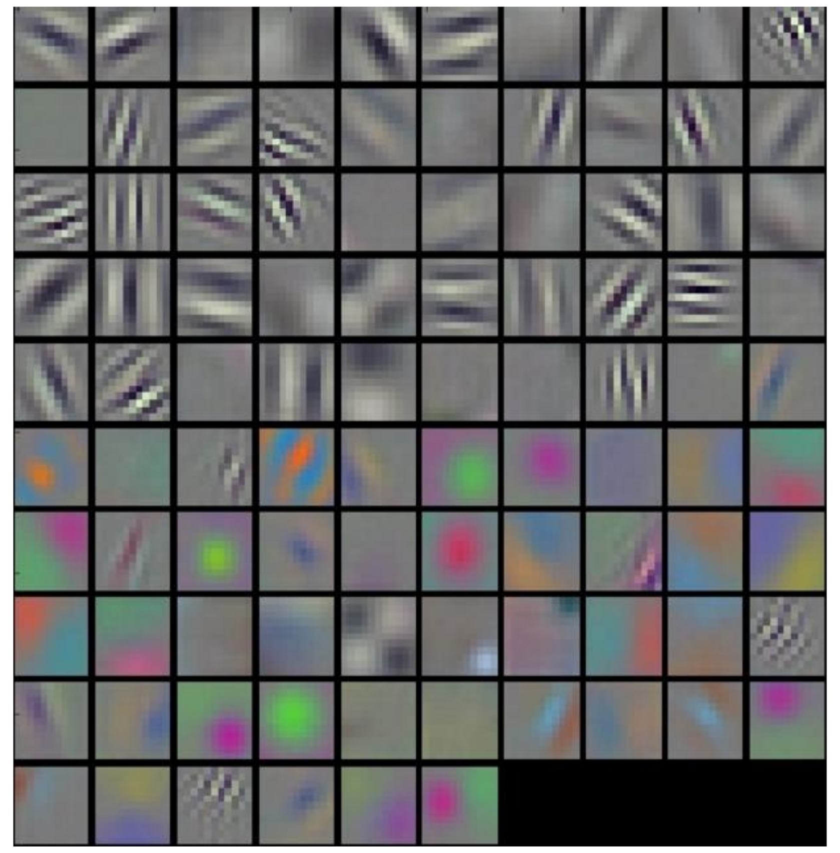
$$\mathcal{H}(\boldsymbol{x}_i^t, \boldsymbol{x}_j^s) = \sum_{h=1}^{D} w_h[\kappa(\phi_h^{i,t}, \phi_h^{j,s}) + \kappa(\phi_h^{j,s}, \phi_h^{i,t})]$$

- **Increase the Number of Nearest Neighbors for Hard Samples Iteratively**

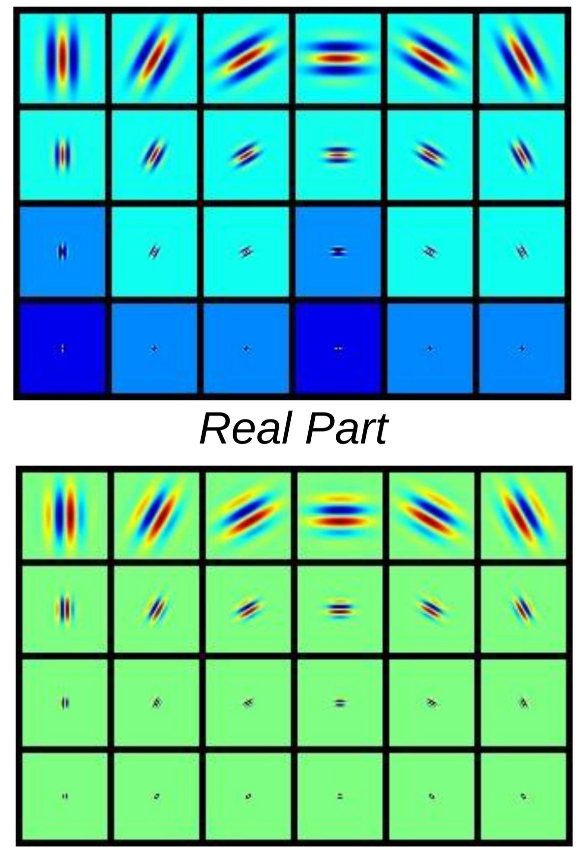
$$\mathcal{K}_i^{m+1} = \begin{cases} \mathcal{K}_i^m + \sigma_0, & \hat{y}_i^t \neq y_i^t \\ \mathcal{K}_i^m + \sigma_1, & \hat{y}_i^t = y_i^t \ \ and \ \ \mathcal{H}_i^m \geq \delta \\ \mathcal{K}_i^m, & \hat{y}_i^t = y_i^t \ \ and \ \ \mathcal{H}_i^m < \delta \end{cases}$$

Pipeline of Selective Joint Fine-Tuning :





Filters in the First Conv Layer of AlexNet



Gabor Filters

Experiments:

- **Codes and Models of Selective Joint Fine-Tuning**
  https://github.com/ZYYSzj/Selective-Joint-Fine-tuning
- **Fine-Grained Object Recognition** (Stanford Dogs 120).

Method	mean Class Acc (%)
HAR-CNN	49.4
Local Alignment	57.0
Multi Scale Metric Learning	70.3
MagNet	75.1
Web Data + Original Data	85.9
Training from scratch using target domain only	53.8
Selective joint training from scratch	83.4
Fine-tuning w/o source domain	80.4
Joint fine-tuning with all source samples	85.6
Selective joint FT with random source samples	85.5
Selective joint FT w/o adaptive NN retrieval	88.3
Selective joint FT with Gabor filter bank	87.5
Selective joint fine-tuning	90.2
Selective joint FT with Model Fusion	90.3

- **Scene Recognition** (MIT Indoor 67).

Method	mean Class Acc (%)
MetaObject-CNN	78.9
MPP + DFSL	80.8
VGG-19 + FV	81.0
VGG-19 + GoogleNet	84.7
Multi scale + multi model ensemble	86.0
Fine-tuning w/o source domain	81.7
Selective joint FT with ImageNet (i)	82.8
Selective joint FT with Places (ii)	85.8
Selective joint FT with hybrid data (iii)	85.5
Average the output of (ii) and (iii)	86.9